AD_____

Award Number: W81XWH-09-1-0694

TITLE: Assessing the role of copy number variants in prostate cancer risk and progression using a novel genome-wide screening method

PRINCIPAL INVESTIGATOR: Dr. Donna Lehman

CONTRACTING ORGANIZATION: University of Texas Health Science Center
San Antonio, TX 78229

REPORT DATE: October 2011

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-10-2011	Annual	15 Sep 2010 - 14 Sep 2011

4. TITLE AND SUBTITLE
Assessing the role of copy number variants in prostate cancer risk and progression using a novel genome-wide screening method

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-1-0694

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Donna Lehman

E-Mail: lehman@uthscsa.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Texas Health Science Center
San Antonio, TX 78229

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Individual copy number variations in the genome may play a substantial role in influencing trait variation, yet due to technical limitations they have been understudied. We have performed the first genome-wide association of copy number variants and risk for prostate cancer in Mexican Americans. We found a highly protective deletion on 8q24 which is present in Mexican Americans but extremely rare in Caucasians. Due to the strong effect of this deletion, this discovery has implications for prostate cancer risk assessment and for understanding the etiology of prostate cancer. This variant warrants further study. We have also confirmed a deletion on 2p24 to be associated with risk for aggressive prostate cancer in non-Hispanic Caucasians and have shown that this allele is very rare in Mexican Americans and therefore not an influential factor in this population. This supports our hypothesis that heritable structural variation may affect risk for prostate cancer and/or its progression. Moreover, these variants may be unique to an ethnic population and underscores the need to investigate genetic risk in multiple populations. As genes are identified from these studies, they may prove to be both useful biomarkers for early diagnosis and/or excellent therapeutic targets for both prevention and treatment of prostate cancer.

15. SUBJECT TERMS
Heritable copy number variation, prostate cancer, Hispanic

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	10	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

Prostate cancer is known to have a strong genetic component. Thus, the identification of the heritable genetic alteration(s) that precedes or increases susceptibility to somatic cancerous changes in the prostate could likely lead to improved identification of high risk individuals for early screening and possibly to new treatment strategies. Standard methodologies, including linkage analysis in familial prostate cancer patients and genome-wide single nucleotide polymorphism (SNP) screening have not identified sufficient genetic alterations to account for the hereditary component of prostate cancer. Recently, it has become apparent that structural variation comprises similar diversity of human genomes as SNPs and may play a significant role in disease susceptibility and resistance. Since CNV regions often contain genes, parts of genes, or regulatory regions, they could result in different levels of gene expression. In addition, through deletion or insertion of stretches of DNA sequence, CNVs may alter the local genomic architecture resulting in differences in the epigenome. Thus, they may play a substantial role in influencing trait variation, yet due to technical limitations they have been understudied, and little is known about this new class of variant, including their distribution in most human populations and impact on common diseases. The goal of the current research is to screen the entire autosomal genome for these variants in constitutional DNA to assess their role in risk of development of prostate cancer and then evaluate any direct effect on the prostate.

BODY

Our closing comments from the "Conclusions" section of our last progress report were: Over this first project period we have gained an immense amount of experience in the complex area of genome-wide copy number variant identification and analyses. Although there is still no complete consensus on statistical analytical methods for determining copy number calls from SNP arrays, much progress has been made and 2 software programs have become the predominant choice. These are PennCNV[1] and QuantiSNP[2]. In another ongoing study of diabetes, we have applied these programs and their various tools to data from much denser Illumina arrays (500K – 1M duo arrays) for approximately 1200 Mexican American subjects from a local family-based cohort. We have been able to clearly distinguish breakpoints of CNVs due to the dense spacing of markers on these arrays. This has enabled better design of qPCR assays for confirmatory analyses. It also provides much greater confidence in copy calls allowing the detection of rare variants as well as those that are small. Given our current experience with very dense SNP arrays, we see a need for applying these methods to this prostate cancer project. A comparison of 100 additional cases with 100 "hyper-normal" controls (i.e., older age) that are better matched for admixture using Illumina's OmniExpress array, which consists of approximately 750,000 probes, and the new statistical methodology will increase our ability to identify and better define copy variable regions, and importantly, those that are rare and/or small. We can then follow up statistically associated regions in fewer samples to validate. We anticipate a better assessment of copy variants to test for global burden as well as pathway analysis.

In this year of the project, we have completed the above recommended strategy. We selected cases with the earliest age-at-onset of PCa and controls that were among the oldest

controls. The samples were run on the Illumina OmniExpress array and the data analyzed using the programs PennCNV[1] and QuantiSNP[2]. We genotyped 192 samples total: 96 cases, 92 controls, and 4 replicate samples. All 4 replicate samples had SNP genotyping reproducibility rates >0.9999. All samples had SNP genotyping call rates >0.99. Three samples (2 cases and 1 control) did not meet criteria for CNV calling due to LogR ratios having standard deviation >0.3 (a standard QC setting in the program PennCNV). Therefore, for association testing in CNVtools there were 94 cases and 91 controls. The average age of cases genotyped on the OmniExpress was 60.43 ± 6.33 years and the average age of controls was 70.88 ± 5.89 years. After QC criteria, the average age of cases used for association testing with CNVtools was 60.36 ± 6.36, and the average age for controls was 70.89 ± 5.93. The groups did not differ in admixture estimates based upon the individual measures that were previously calculated for this cohort[3] as shown in Table 1.

Table 1. Admixture estimates of cases and controls

Admixture estimate	Cases	Controls	P-value
% European American	0.587 ± 0.180	0.615 ± 0.182	0.38
% Native American	0.385 ± 0.185	0.358 ± 0.189	0.41
% African American	0.028 ± 0.040	0.027 ± 0.039	0.82

We identified 462 unique CNV regions that were detected in at least 2 subjects. Using the combined genotyping and likelihood ratio association testing model implemented in CNVtools,, we genotyped each individual and performed association testing for all 462 CNV regions. We observed significant association ($p<0.05$) with 8 CNVs. 2 of these 8 were significant after Bonferroni correction.

We next validated 7 associated CNVs in the full dataset of 630 Mexican American subjects using quantitative PCR. One CNV we have been unable to design real-time PCR primer for due to its location in a complex region of the genome. Real time PCR data was analyzed using CopyCaller software (Applied Biosystems, Valencia CA), which implements the $\Delta\Delta$ct method. We identified integer copy number calls by plotting histograms of the raw calculated copy number calls. For 5 of the CNVs non-overlapping Gaussian distributions representing integer copy number calls were clearly present as shown in Figure 1. For 2 CNVs, we could not cluster and chose to analyze them based on the quantitative value of the calculated call. We again tested for association with PCa using logistic regression, and for those with discrete calls, we also used Fisher's exact test. The results from Fisher's test were consistent with those from logistic regression. The results of analyses are shown below in Table 2.

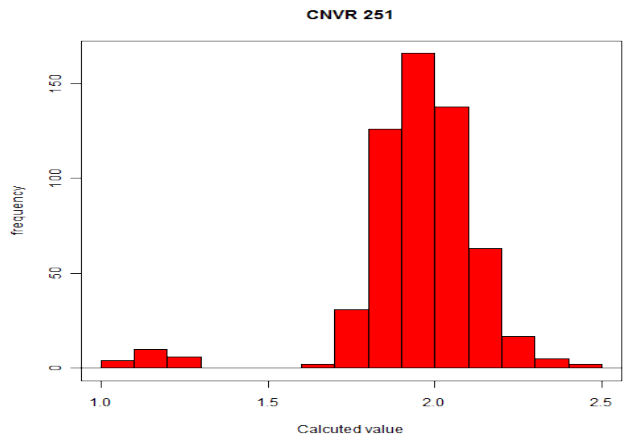

Figure 1. Genotyping of CNV in 630 subjects. Histogram of calculated copy number values for CNVR251. Non-overlapping Gaussian distributions are present representing integer copy number values of 1 and 2.

Table 2. Association of CNVs in 630 SABOR Mexican-American subjects

					No. bearing variant					
CNV	**Chr**	**Start (bp)**	**End (bp)**	**Candidate gene**	**Cases (n=190)**	**Controls (n=440)**	**OR**	**OR 95% CI Lower**	**OR 95% CI Upper**	**LR. P-value**
CNVR251	8	135062170	135065947	NDRG1	1	19	0.101	0.764	0.013	0.026
CNVR158	5	113150200	113171928	MCC, APC	9	15	1.487	3.499	0.632	0.363
CNVR186	6	69687698	69690567	BAI3	7	30	0.671	1.433	0.314	0.302
CNVR426	18	67209141	67217271	DOK6	4	20	0.484	1.454	0.161	0.196
CNVR88	3	89402447	89417171	EPHA3	12	41	0.652	1.284	0.332	0.216
CNVR54	2	130702757	130731130	RAB6C	1.99*	1.99*	1.069	3.205	0.357	0.905
CNVR156	5	106229524	106230898	EFNA5	1.96*	1.99*	1.959	4.259	0.901	0.090

*The mean of the calculated quantitative value for the group, where "2" would be interpreted as the standard copy number of 2 alleles.

The most striking result was that of CNVR251. This CNV appeared to be a deletion that is much more prevalent in the elderly controls than in the cases (p=0.011, Fisher's test). To further explore this variant and confirm the deletion, we designed primers flanking and within the deleted genomic region (as defined by results reported in the Database of Genomic Variants from high density arrayCGH). We tested DNA from 6 subjects bearing the variant using the flanking primers and conducted direct sequencing of the PCR products. The sequence data showed that all 5 control samples had identical breakpoints on chromosome 8q24 of 8486 base pairs. PCR amplification of the DNA from the case bearing the deletion resulted in a band of similar size; however, we have not been successful in sequencing this sample to date. We are continuing to pursue verification by sequencing. Only 3 of 1530 Caucasian subjects carried the deletion, indicating this deletion is not likely to affect risk in the Caucasian population. In order to

facilitate accurate and affordable confirmation of this CNV in other cohorts we sought to identify nearby SNPs which could be used to impute this deletion. However, upon examining the linkage disequilibrium within this region we were unable to identify any SNPs with $R^2 > 0.4$ with this deletion, and concluded that this CNVR was not imputable (Figure 2). We are currently working on a PCR based genotyping assay with discrete results using these flanking primers and a nested primer. This approach will facilitate accurate and affordable genotyping of this CNV by other groups wishing to validate our finding.

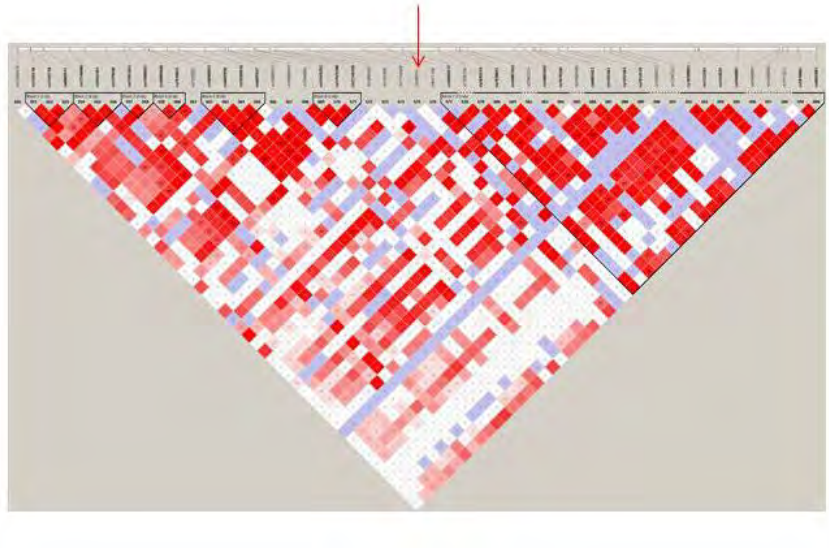

Figure 2. Plot of LD (AbsD' measure) determined by the program Haploview using genotypic information from 188 SABOR/PREF Hispanic samples genotyped on the Illumina OmniExpress array. R^2 values are shown inside squares for each marker comparison. Arrow marks position of CNVR251

The deletion is located ~6.3 Mb from the Myc gene locus and is therefore likely distinct. Observation of this region using the UCSC genome browser revealed that the deleted sequence contains a conserved transcription factor binding site for NKX3-1, an androgen regulated homeobox gene involved in prostate development.[4] This deletion may therefore affect expression of a nearby gene related to prostate proliferation.

Task 2 of our SOW is to determine the effect of identified CNVs on expression in lymphoblastoid cell lines as a surrogate tissue. We are growing cell lines for all control subjects bearing variants and corresponding control subjects without the variant for CNVRs 251, 426, and 186. RNA will be isolated and cDNA made from each cell line. We have designed RT-PCR assays for genes flanking each CNVR. We will be conducting expression analyses using these assays when they arrive. In addition, we also have access to stored peripheral blood mononuclear cells (PBMCs) for several of the subjects bearing these variants. We will isolate RNA from these samples as well and test for differential gene expression at these loci.

Task 3 of our SOW is directed at determining whether germ-line CNV loci are additional targets of aneuploidy in tumors. The recent large scale genome sequencing data of the Cancer Genome Atlas may allow us to conduct this task *in silico*. We are currently exploring this possibility as well as identifying other available datasets. These samples may not reflect those of Mexican American origin, however, and we may need to assay available prostatectomy samples from the San Antonio tissue bank.

Replication of a heritable deletion and risk for aggressive PCa

In addition to performing CNV discovery in the SABOR cohort, we are also utilizing this resource to examine and validate CNVs that have been reported by other research groups. Liu et al.[5] reported a deletion on 2p24.3 associated with aggressive prostate cancer in a non-Hispanic Caucasian population. We tested whether this finding could be confirmed both in non-Hispanic and Hispanic Caucasians from the SABOR and PREF. Among non-Hispanic Caucasians, carrying a homozygous deletion was significantly associated with aggressive prostate cancer as defined by a Gleason sum ≥8 [odds ratio (OR), 27.99; 95% confidence interval (CI), 1.99-392.6; P = 0.007], and carrying either a homozygous or heterozygous deletion was suggestive of an association with overall prostate cancer risk (OR, 1.37; 95% CI, 0.93-2.00; P = 0.09). Using a one-side p-value from logistic regression based on the *a priori* knowledge that this deletion was association, this deletion was associated with overall prostate cancer risk (P = 0.03). Among Hispanic Caucasians, this deletion is much less prevalent (minor allele frequencies of 0.059 and 0.024 in non-Hispanic and Hispanic Caucasians respectively) and was not associated with risk for prostate cancer (OR, 0.99; 95% CI, 0.39-2.31; P = 1). No aggressive Hispanic Caucasian cases carried this germ-line deletion. This study independently confirmed the first germ-line CNV to be associated with risk for aggressive prostate cancer in non-Hispanic Caucasians. However, we found a lack of evidence for the role of this CNV in risk for prostate cancer in Hispanic Caucasians. A manuscript reporting these results has been submitted to the journal *Cancer Epidemiology Biomarkers and Prevention*.

KEY RESEARCH ACCOMPLISHMENTS

- We have performed a genome-wide screen of CNVs using dense SNP arrays and improved statistical techniques in 100 Mexican American cases with earliest age at onset and 100 Mexican American hyper-normal controls matched on admixture.
- We have discovered a low frequency germ-line deletion that is unique to Mexican Americans. Carriers of this deletion appear to be at significantly reduced risk for PCa (OR 0.1).
- We have independently confirmed the first germ-line CNV to be associated with risk for aggressive prostate cancer in non-Hispanic Caucasians at chromosome 2p24.3 and have shown that this allele is very rare in Mexican Americans and therefore not an influential factor in this population.

REPORTABLE OUTCOMES

- Blackburn A, Gelfond J, Yao L, Dean A, Hernandez J, Thompson IA, Leach RJ, Lehman DM (2011) Risk for aggressive prostate cancer and a heritable deletion at 2p24.3 in non-Hispanic and Hispanic Caucasians. Cancer Epi Bio Prev (submitted).
- Blackburn A, Gelfond J, Yao L, Thompson IA, Leach RJ, Lehman DM (2011). A heritable deletion on 8q24 lowers risk for prostate cancer in Mexican Americans. Abstract to be presented at the Cancer Therapy and Research Center Annual Symposium, UT Health Science Center, San Antonio TX

CONCLUSION

We have performed the first genome wide association of copy number variants and risk for prostate cancer in Mexican Americans. We found a highly protective deletion on 8q24 which is present in Mexican Americans but extremely rare in Caucasians. Due to the strong effect of this deletion, this discovery has implications for prostate cancer risk assessment and for understanding the etiology of prostate cancer. This variant warrants further study. We have also confirmed a deletion on 2p24 to be associated with risk for aggressive prostate cancer in non-Hispanic Caucasians and have shown that this allele is very rare in Mexican Americans and therefore not an influential factor in this population. This supports our hypothesis that heritable structural variation may affect risk for PCa and/or its progression. Moreover, these variants may be unique to ethnic population and underscores the need to investigate genetic risk in multiple populations. As genes are identified from these studies, they may prove to be both useful biomarkers for early diagnosis and/or excellent therapeutic targets for both prevention and treatment of prostate cancer.

REFERENCES

(1) Wang K, Li M, Hadley D et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007 November;17(11):1665-74.

(2) Colella S, Yau C, Taylor JM et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007;35(6):2013-25.

(3) Beuten J, Halder I, Fowler SP et al. Wide disparity in genetic admixture among Mexican Americans from San Antonio, TX. *Ann Hum Genet* 2011 July;75(4):529-38.

(4) Kojima C, Zhang Y, Zimmer WE. Intronic DNA elements regulate androgen-dependent expression of the murine Nkx3.1 gene. *Gene Expr* 2010;15(2):89-102.

(5) Liu W, Sun J, Li G et al. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res* 2009 March 15;69(6):2176-9.

BIBLIOGRAPHY

1. Blackburn A., Gelfond J., Goring H.H., Beuten Y., Thompson I., Leach RJ, and Lehman DM. (2009) Identification of Copy Number Variable Regions (CNVRs) Associated with Risk of Prostate Cancer in Mexican-Americans. Abstract presented at 59th Annual meeting of the American Society of Human Genetics, Honolulu HI, October 2009
2. Lehman DM. Identification of Copy Number Variable Regions (CNVRs) Associated with Risk of Prostate Cancer in Mexican-Americans, DOD IMPaCT conference, Health Disparities, Department of Defense, Orlando, F,L March 2011 (Invited Speaker)
3. Blackburn A, Gelfond J, Yao L, Dean A, Hernandez J, Thompson IA, Leach RJ, Lehman DM (2011) Risk for aggressive prostate cancer and a heritable deletion at 2p24.3 in non-Hispanic and Hispanic Caucasians. Cancer Epi Bio Prev (submitted).
4. Blackburn A, Gelfond J, Yao L, Thompson IA, Leach RJ, Lehman DM (2011). A heritable deletion on 8q24 lowers risk for prostate cancer in Mexican Americans. Abstract to be presented at the Cancer Therapy and Research Center Annual Symposium, UT Health Science Center, San Antonio TX

Personnel: Donna Lehman, Robin Leach, Jon Gelfond, August Blackburn